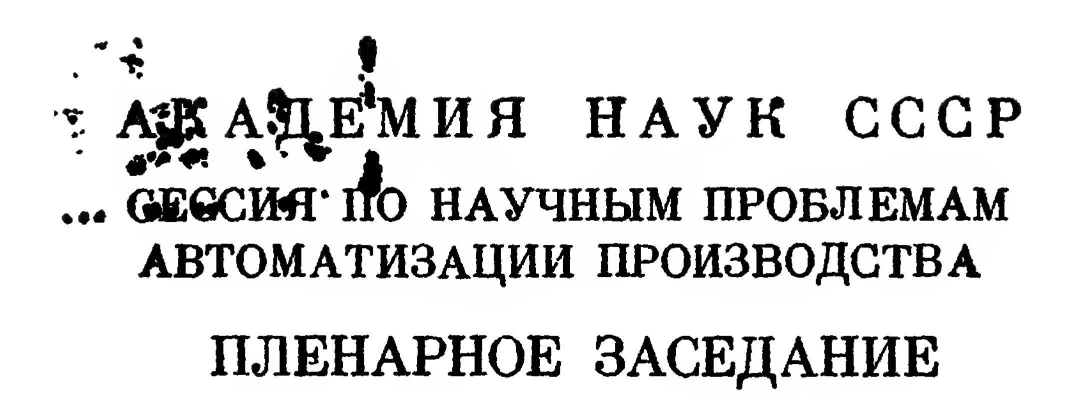
АКАДЕМИЯ НАУК СССР
СЕССИЯ ПО НАУЧНЫМ ПРОБЛЕМАМ
АВТОМАТИЗАЦИИ ПРОИЗВОДСТВА
ПЛЕНАРНОЕ ЗАСЕДАНИЕ

А. Н. КОЛМОГОРОВ

ТЕОРИЯ ПЕРЕДАЧИ ИНФОРМАЦИИ



ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР
Москва 1956



АКАДЕМИЯ НАУК СССР

СЕССИЯ ПО НАУЧНЫМ ПРОБЛЕМАМ АВТОМАТИЗАЦИИ ПРОИЗВОДСТВА

ПЛЕНАРНОЕ ЗАСЕДАНИЕ

А. Н. КОЛМОГОРОВ

ТЕОРИЯ ПЕРЕДАЧИ ИНФОРМАЦИИ



ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР

Москва 1956

I. ВОЗНИКНОВЕНИЕ ТЕОРИИ И ЕЕ СОДЕРЖАНИЕ

Выступая второй раз перед общим собранием нашей Академии, мне хочется начать с замечания, что тема моего сегодняшнего сообщения тесно связана с темой доклада «Статистическая теория колебаний с непрерывным спектром», который я делал на общем собрании Академии в 1947 году *. В самом деле, наиболее богатая содержанием и наиболее увлекательная с чисто математической стороны часть теории информации — теория стационарно работающих каналов связи, передающих непрерывные сообщения, — не могла бы быть создана без заранее разработанной теории стационарных случайных процессов и, в частности, без спектральной теории таких процессов.

Интерес к разнообразным задачам передачи и хранения информации имеет большую давность. Давно, по существу, возникали и вопросы об оценке «количества» информации. Вопрос о возможности введения универсальной числовой меры для количества информации особенно важен в случаях необходимости преобразования информации одного рода в информацию качественно другого рода. Типичной задачей на преобразование информации является задача табулирования непрерывных функций от непрерывно меняющихся аргументов. Если, например, функцию $f(x,y)$ от двух аргументов

$$0 \leqslant x \leqslant 1, \; 0 \leqslant y \leqslant 1$$

необходимо задать с точностью до ε и известно, что ее приращение Δf не превосходит по абсолютной величине ε, когда

$$|\Delta x| \leqslant \varepsilon_x, \; |\Delta y| \leqslant \varepsilon_y,$$

* См. [1]. Более подробное изложение математической стороны вопросов, которым был посвящен этот мой доклад в 1947 году, выделено в статью [2]. Сейчас аналогичную роль играет глава II настоящей брошюры, воспроизводящая с небольшими изменениями доклад, прочитанный Б. В. Гнеденко от моего имени на симпозиуме по теории информации Американского института радиоинженеров (Массачузетский технологический институт, сентябрь 1956) и основанный во многих частях на докладе, прочитанном мною совместно с И. М. Гельфандом и А. М. Ягломом на Всесоюзном математическом съезде в июне 1956 г.

то заведомо достаточно табулировать функцию f с шагом ε_x по x и ε_y по y, т. е. задать приблизительно

$$N \backsim \frac{1}{\varepsilon_x \varepsilon_y}$$

ее значений. Если допустить, что $|f| \leqslant 1$, то для фиксации одного значения функции достаточно приблизительно

$$K \backsim \log_{10}\left(\frac{1}{\varepsilon}\right)$$

десятичных знаков. Таким образом, если следовать намеченному сейчас плану, всего в таблицу придется ввести приблизительно

$$NK \backsim \frac{1}{\varepsilon_x \varepsilon_y} \log_{10}\left(\frac{1}{\varepsilon}\right)$$

десятичных знаков (не считая записи значений аргументов x и y, которая стандартна для всех функций f, подчиненных сделанным ограничениям). Общеизвестно, что такой примитивный подход к делу, особенно в случае функций нескольких переменных, привел бы к таблицам огромного объема и практически неосуществимым. На самом деле при табулировании достаточно «гладких» функций выбирают шаг таблицы по независимым переменным значительно бо́льшим, а для нахождения значений функции при промежуточных значениях аргумента прибегают к интерполированию с помощью разностей до какого-либо порядка p (в случае таблиц функций нескольких переменных часто до четвертого, или шестого порядка). Кроме того, часто первые знаки $f(x, y)$, повторяющиеся для близких значений аргументов, не выписывают в каждой клеточке таблицы, соответствующей определенной паре значений (x, y), а выписывают, например, лишь в начале строки, состоящей из многих таких клеточек.

При всей общеизвестности и давнем происхождении описанных приемов построения таблиц соответствующие общие теоретические исследования, которые должны выяснить *минимальное* количество информации, необходимое для фиксации с заданной точностью ε произвольной функции f, подчиненной лишь тем или иным общим условиям, находятся еще в начальной стадии. Например, только недавно в моей заметке [3] была явно дана формула *

$$H_\varepsilon(F^n_{p,\alpha}) \asymp \left(\frac{1}{\varepsilon}\right)^{\frac{n}{p+\alpha}}, \quad (1)$$

* Формулу эту можно назвать формулой А. Г. Витушкина, так как в его работе [4] по другому поводу выяснена роль показателя $\frac{n}{p+\alpha}$ и,

указывающая порядок роста при $\varepsilon \to 0$ количества информации, необходимого для фиксации заданной в ограниченной области функции n переменных с ограниченными производными первых p порядков и производными порядка p, удовлетворяющими условию Липшица степени α.

Вполне понятно, что вопросы преобразования и хранения информации имеют решающее значение при разработке конструкций и способов употребления современных вычислительных машин. Здесь «объем памяти» запоминающего устройства характеризуется числом удерживаемых этим устройством двоичных знаков (0 или 1). Способ измерения количества информации при помощи сравнения любой информации с информацией в виде некоторого числа двоичных знаков стал сейчас стандартным в теории информации.

Переходя к задачам из теории «каналов связи» в той ее форме, которая связана с работой телеграфных или телефонных линий и подобных им устройств, начнем с очень простой задачи о передаче последовательности десятичных знаков

$$a_1, a_2, \ldots, a_n, \ldots \tag{2}$$

при помощи последовательности двоичных знаков

$$b_1, b_2, \ldots, b_n, \ldots \tag{3}$$

Допустим, что последовательность (2) возникает во времени по одному знаку в единицу времени. Спрашивается, с какой скоростью надо передавать двоичные знаки последовательности (3), если желать, чтобы восстановление по последовательности (3) исходной последовательности (1) осуществлялось без неограниченно возрастающего со временем запоздания. Если изображать каждый знак a_n отдельно двоичными знаками, то на один знак a_n приходится тратить четыре двоичных знака (так как $2^3 \leqslant 8 < 10$ и только $2^4 = 16 > 10$). Если передавать каждую пару (a_{2n-1}, a_{2n}) знаков последовательности (2) группой двоичных знаков, то на нее придется тратить семь двоичных знаков (так как $2^6 = 64 < 100$ и только $2^7 = 128 > 100$), для передачи трех десятичных знаков достаточно тратить десять двоичных знаков (так как $2^9 < 1000$, но $2^{10} = 1024 > 1000$) и т. д. Соответствующие этим системам передачи скорости образования знаков последовательности (3), т. е. соответствующее число двоичных знаков, которые необходимо в среднем передавать

по существу, содержится половина доказательства формулы (1), а именно, доказательство того факта, что порядок $H_\varepsilon(F^n_{p,\alpha})$ не может быть меньше указываемого формулой (1).

за единицу времени, равны

$$v_1 = 4;\ v_2 = \frac{7}{2} = 3{,}5;\ v_3 = \frac{10}{3} = 3{,}33\ldots$$

Легко доказать, что возникающий таким образом ряд чисел сходится к пределу

$$v = \log_2 10 = 3{,}32\ldots$$

Именно это v и является нижней гранью тех скоростей создания знаков последовательности (3), при которых можно без систематически возрастающего запоздания передавать последовательность (2).

Уже на этом нарочито наивном примере можно наблюдать многие типичные явления, открытые современной теорией передачи информации в применении к значительно более общим видам каналов связи:

1) скорость создания информации, несмотря на разнородность ее качества, допускает разумное количественное измерение. В частности, мерой скорости создания информации при фиксировании в единицу времени в среднем v знаков, каждый из которых может принимать k значений, разумно считать число

$$\overline{H} = v \log_2 k.$$

(В нашем случае

$$\overline{H} = \log_2 10 = 3{,}32\ldots,$$

для последовательности (2) и

$$\overline{H}' = v \log_2 2 = v$$

для последовательности (3) при создании v знаков в единицу времени).

Для дальнейшего заметим, что в случае точно (без ошибок) работающих каналов связи скорость выдаваемой ими информации разумно называть их «пропускной способностью». В нашем случае пропускная способность последовательности (3), рассматриваемой как канал связи, равна $\overline{C} = \overline{H}' = v$. В общем случае пропуская способность определяется несколько иначе, что будет позднее объяснено. После этого замечания сформулируем второй общий принцип:

2) для возможности передачи возникающей со скоростью $\overline{H}$ информации через канал связи с пропускной способностью $\overline{C}$ без систематически возрастающего запоздания необходимо условие

$$\overline{H} \leqslant \overline{C}.$$

Наиболее существенен, однако, третий принцип, содержание которого в приводимой сейчас формулировке еще очень расплывчато, но в котором и заключается, пожалуй, наиболее существенное новое, внесенное в теорию передачи информации по каналам связи работами Шеннона:

3) даже при значительной качественной разнородности информации, подлежащей передаче, и той информации, которую способен непосредственно воспринимать и выдавать канал связи, в случае

$$\bar{H} < \bar{C},$$

в принципе возможно производить передачу без систематически возрастающего запаздывания, если подлежащая передаче информация перед поступлением в канал связи надлежащим образом «кодируется», а после выхода из канала «декодируется». Однако, вообще говоря, при приближении $\bar{H}$ к $\bar{C}$ способы кодирования и декодирования неизбежно усложняются и возникает все большее и большее (при $\bar{H} \to \bar{C}$) запаздывание передачи.

Заметим, что запаздывание, о котором идет речь в принципе 3), возрастает при $\bar{H} \to \bar{C}$, но при постоянном источнике информации с $\bar{H} < \bar{C}$ может быть оценено некоторой величиной τ, которая сохраняется при неограниченном продолжении работы канала.

В виде следующего примера рассмотрим передачу текста, состоящего из букв русского алфавита. Так как различных букв в нашем алфавите 33, то можно формально образовать

$$N = 33^n$$

различных «текстов» длины n (т. е. последовательностей из n букв). Количество информации, содержащейся в указании одного определенного такого текста, равно, по-предыдущему,

$$I = n \log 33$$

(здесь и всюду далее мы употребляем логарифмы при основании 2, не оговаривая этого специально). Существование стенографии показывает, однако, что реальные языковые тексты можно передавать значительно более кратким образом. Это вполне понятно: число N^* «осмысленных» текстов из n букв, конечно, несравненно меньше N. Поэтому, в принципе возможна система записи («идеальная» стенография), которая требовала бы для фиксации любого из этих N^* осмысленных текстов лишь

$$I^* \sim \log N^*.$$

двоичных знаков. Полное использование этой принципиальной возможности «сжатия текста», конечно, лежит за пределами осуществимого при помощи каких-либо систем стенографирования или кодирования текста, следующих достаточно простым формальным правилам: закономерности построения реального языкового текста вряд ли когда-либо будут полностью формализованы.

Мы обратим сейчас внимание лишь на одну особенность языковых текстов: различные буквы встречаются в них с различной частотой. Если фиксировать числа

$$n_1, n_2, \ldots, n_k$$

вхождений в текст длины n каждой из букв

$$a_1, a_2, \ldots, a_k,$$

(естественно, что при этом

$$n_1 + n_2 + \ldots + n_k = n),$$

то число текстов длины n сократится до

$$N' = \frac{n!}{n_1!\, n_2! \ldots n_k!}$$

Пользуясь так называемой формулой Стирлинга, которая нужна здесь даже лишь в ослабленной форме:

$$\log(n!) \sim n \log n,$$

легко подсчитать, что при больших n_i

$$I' = \log N' \sim -n \sum_i p_i \log p_i,$$

где

$$p_i = \frac{n_i}{n}$$

частоты появления отдельных букв. Полученный результат можно выразить так: при употреблении отдельных букв с частотами p_i количество передаваемой информации «на одну букву текста» равно

$$H = -\sum_i p_i \log p_i. \qquad (4)$$

В случае равных частот

$$p_1 = p_2 = \cdots = p_k = \frac{1}{k}$$

мы вновь получаем

$$H = \log k;$$

при любых других частотах p_i

$$H < \log k.$$

Легко указать приемы, кодирования, позволяющие передавать произвольный текст, в котором буквы $a_1, a_2, \ldots, a_k$ встречаются с частотами $p_1, p_2, \ldots, p_k$, затрачивая в среднем (при передаче достаточно длинных текстов) на одну букву число двоичных знаков, сколь угодно мало отличающееся от H, определенного по формуле (4). Это еще одно проявление третьего из сформулированных выше принципов. Мы оставим, однако, теперь рассмотрение элементарных примеров и перейдем к элементам общей теории.

Формула (4) допускает, кроме чисто статистической (в которой p_i являются частотами), вероятностную интерпретацию. Пусть со всем общим образом дано распределение вероятностей

$$P(\xi = x_i) = p_i \tag{5}$$

случайного объекта ξ. Если дано только это распределение, то в ответе на вопрос о том, какому из возможных значений

$$x_1, x_2, \ldots, x_k$$

равно в действительности ξ, остается некоторая неопределенность. Сообщая, что ξ равно такому-то определенному x_i, мы устраняем эту неопределенность, т. е. сообщаем некоторую дополнительную информацию об объекте ξ. Мерой неопределенности распределения (5) является его энтропия, выражающаяся уже знакомой нам формулой (4). Эта же формула выражает и количество информации, необходимое для устранения неопределенности в задании ξ лишь распределением (5), т. е. информация, содержащаяся в указании точного значения ξ.

Пусть теперь далее дано совместное распределение вероятностей

$$P(\xi = x_i, \eta = y_j) = p_{ij}$$

двух случайных объектов ξ и η. Если $\eta = y_j$ задано, то ξ получает условное распределение

$$P(\xi = x_i \mid \eta = y_j) = p_{i|j}$$

Количество информации, содержащееся в указании точного значения ξ, если значение $\eta = y_j$ уже известно, равно

$$H(\xi \mid \eta = y_j) = -\sum_i p_{i|j} \log p_{i|j},$$

в среднем же оно равно

$$MH(\xi \mid \eta) = -\sum_j P(\eta = y_j) \sum_i p_{i|j} \log p_{i|j}.$$

Естественно считать, что разность

$$I(\eta, \xi) = H(\xi) - MH(\xi \mid \eta) \qquad (6)$$

есть то количество информации относительно ξ, которое уже содержится в задании η. Легко подсчитать, что $I(\eta, \xi)$ можно записать в виде

$$I(\eta, \xi) = \sum_{i,j} p_{ij} \log \frac{p_{ij}}{P(\xi = x_i)\, P(\eta = y_i)}, \qquad (7)$$

где ξ и η играют совершенно одинаковую роль: количество информации в η относительно ξ и количество информации в ξ относительно η численно равны друг другу.

Легко проверить, что всегда

$$I(\xi, \xi) = H(\xi). \qquad (8)$$

Теперь без труда определяются скорость передачи информации через канал связи, работающий с ошибками, и пропускная способность таких каналов (пропускная способность определяется, попрежнему, как верхняя грань возможной скорости передачи информации). Приведу еще один элементарный пример. Пусть «на входе» в канал подаются знаки η, равные 0 или 1, а на выходе получаются соответствующие знаки η', причем

$$P(\eta' = 0) = 1 - \Delta,\ P(\eta' = 1) = \Delta \text{ в случае } \eta = 0,$$
$$P(\eta' = 0) = \Delta,\ P(\eta' = 1) = 1 - \Delta \text{ в случае } \eta = 1.$$

Легко подсчитать, что верхняя грань

$$C = \sup I(\eta, \eta') = 1 + [\Delta \log \Delta + (1 - \Delta) \log (1 - \Delta)] \qquad (9)$$

достигается в случае

$$P(\eta = 0) = P(\eta = 1) = \frac{1}{2}.$$

По формуле (9) количество информации $I(\eta, \eta')$, передаваемое нашим каналом на одну букву, равно заведомо нулю, если

$\Delta=\frac{1}{2}$. Это вполне понятно, так как с вероятностной точки зрения в этом случае знаки η и η' независимы. Максимальное значение пропускной способности $C=1$ получается при $\Delta=0$ и при $\Delta=1$. В случае $\Delta=0$ канал работает без ошибок: с вероятностью единица $\eta=\eta'$. Хотя во втором случае с вероятностью единица $\eta\neq\eta'$, η легко восстанавливается по η' и в этом случае: надо только вместо 0 читать 1 и наоборот. Во всех остальных случаях

$$0<C<1.$$

Когда Шеннон в конце сороковых годов предложил измерять количество информации формулами (4), (6), (7) и обнаружил, что и для каналов связи, работающих с ошибками, при довольно широких предположениях относительно источника информации и строения канала сохраняются сформулированные выше второй и третий принципы, то этим и были созданы основы современной теории информации в виде дисциплины, способной к систематическому развитию.

Во всяком крупном открытии имеются элементы неожиданности. Этим крупное открытие и отличается от постепенно накапливаемых результатов текущей научной работы. Подчеркну здесь, в чем я усматриваю то качественно новое и неожиданное, что содержится уже в описанной элементарной части теории информации.

1. По первоначальному замыслу «информация» не есть скалярная величина. Различные виды информации могут быть чрезвычайно разнообразны. Можно было заранее ожидать, что можно предложить те или иные способы измерять «количество» информации, но было не ясно, существует ли среди этих способов какой-либо один, имеющий принципиальное преимущество перед другими, а главное, было совершенно не ясно, можно ли качественно различные информации, для которых та или иная условная количественная мера одинакова, действительно считать эквивалентными в смысле трудности их передачи через каналы связи, или их хранения в запоминающих устройствах.

Оказалось, что такая «правильная» по преимуществу мера количества информации существует и позволяет решать до конца широкий класс задач, в которых априори независимость решения от более тонких качественных особенностей информации была совершенно не ясна.

2. В случае каналов связи (или запоминающих устройств), работающих с ошибками, можно было опасаться, что достижение передачи с достаточно малой вероятностью ошибки ε связано с очень большим уменьшением скорости передачи. Воп-

реки этим опасениям оказалось, что при условии $\overline{H} < \overline{C}$ сразу становится возможной передача информации со сколь угодно малой вероятностью ошибки. Для пояснения этого обстоятельства укажу на следующее. Пусть дан канал связи с пропускной способностью C. Рассмотрим задачу передачи по нему двоичных знаков, возникающих в числе H за единицу времени.

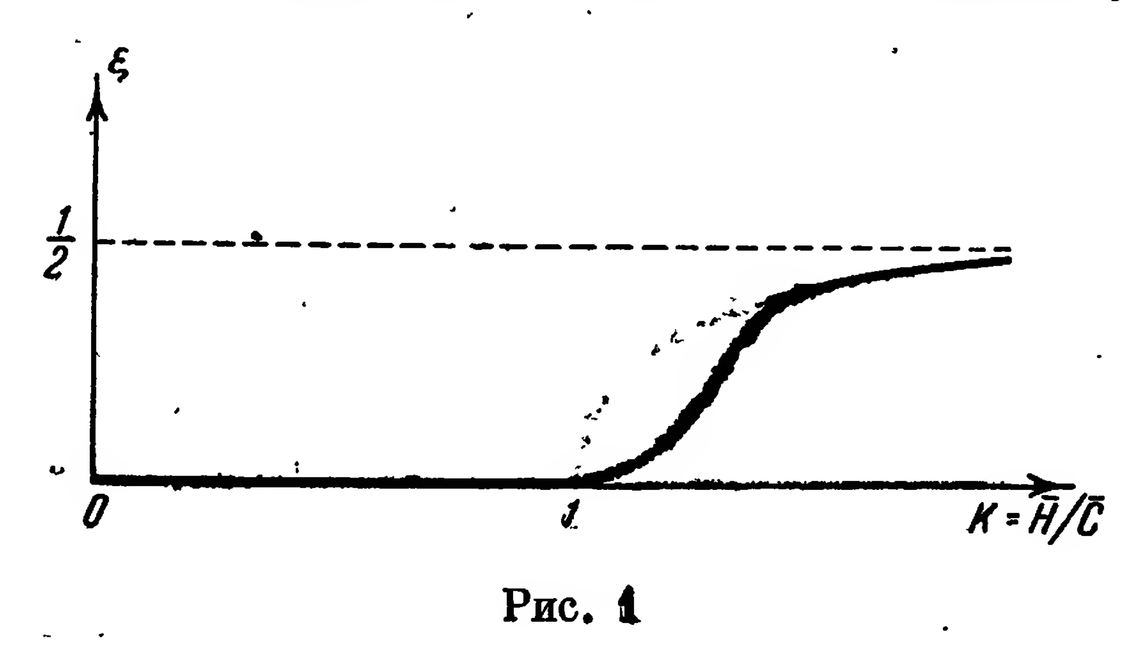


Рис. 1

Тогда достижимый с любым приближением минимум вероятности ошибки в передаче отдельного знака является универсальной функцией $\varepsilon(K)$ отношения $K = \overline{H}/\overline{C}$ (изображенной на рис. 1). При $K \leqslant 1$ она равна нулю, а при дальнейшем возрастании K начинает возрастать и при $K \to \infty$ стремится к половине (аналитически при $K > 1$ функция $\varepsilon(K)$ определяется из уравнения

$$\frac{1}{k} = 1 + \varepsilon \log \varepsilon + (1 - \varepsilon) \log (1 - \varepsilon), \quad \varepsilon < \frac{1}{2}).$$

Таким образом, от значений, точно равных нулю при $K \leqslant 1$, наша функция при $K > 1$ сразу переходит к возрастанию.

Правда, как уже отмечалось ранее, при $\overline{H}$, близких к $\overline{C}$, достижение малой вероятности ошибки требует часто очень сложных правил кодирования информации и приводит к большому запаздыванию ее выдачи на выходе канала. Но именно эта большая сила сложных способов кодирования в смысле устранения ошибок и является принципиально важным выводом, к возможным применениям которого мы еще вернемся.

Обоснование предложенных Шенноном способов измерения количества информации, т. е. формул (4) и (7), можно мыслить двояко. Прежде всего можно выдвинуть те или иные естественные аксиоматические требования, которым мера количества информации должна удовлетворять. В применении к величине $H(\xi)$, т. е. в соответствии с формулой (8) к количеству информации, содержащейся в указании случайного объекта ξ относительно самого этого объекта ξ, образцово ясное изложение

такого подхода к задаче обоснования теории информации дано в работе А. Я. Хинчина [5]. Представляло бы некоторый интерес дать столь же естественную аксиоматику непосредственно более общего понятия количества информации $I(\xi, \eta)$ в одном случайном объекте относительно другого случайного объекта. Дело в том (ср. главу II этого доклада), что в непрерывном случае обычно $H(\xi)$ и $H(\eta)$, взятые отдельно, бесконечны, и $I(\xi, \eta)$ не может вычисляться по формуле (6).

В качестве итога исследований аксиоматического направления можно считать выясненным, что никакой другой столь же естественной скалярной сводной характеристики информации, содержащейся в одном случайном объекте ξ относительно другого случайного объекта η, кроме $I(\xi, \eta)$, существовать не может. Однако, так как «информация» по своей природе не обязана быть (и в действительности не является!) скалярной величиной, то никакие аксиоматические исследования указанного направления не могут ответить на вопрос о том, сколь полно характеризует величина $I(\xi, \eta)$ интересующую нас «информацию». Некоторый ответ на этот вопрос, как уже можно было понять из предшествующего изложения, дают теоремы, обосновывающие второй и третий сформулированные выше общие принципы теории передачи информации. В силу этих теорем в весьма широком классе случаев приспособленность канала связи, или запоминающего устройства к передаче, или хранению информации с достаточной полнотой характеризуется одним единственным числом C (при расчете на единицу времени $\overline{C}$), а приспособленность самой информации к передаче, или хранению — одним единственным числом H (или при расчете на единицу времени $\overline{H}$). Качественное своеобразие информации оказывается при этом не существенным. Трудно переоценить значение такого рода результатов, если учесть широту применения в современной технике таких преобразований, как преобразование изображений, передаваемых в телевидении, в электрические колебания и обратно (в терминологии теории передачи информации это частный случай «кодирования» и «декодирования») и т. п. Вместе с тем надо понимать, что при всей увлекательности идей теории информации подобное стирание качественных особенностей информации имеет место только с известным приближением и при определенных условиях. Эти условия в хорошо исследованных случаях, грубо говоря, состоят в накоплении большого количества однородной информации, в действительной осуществимости сложных методов кодирования и в безвредности возникающего при этом запоздания выдачи информации. Кроме того, следует ясно представлять себе, что точное математическое обоснование

реальности этих представлений пока выполнено лишь при довольно ограниченных предположениях. При необычайном богатстве идей, данных в работах самого Шеннона, изложение в них обычно крайне туманно. Лишь позднее в ряде работ чистых математиков для случая стационарно работающих каналов, передающих дискретные сигналы, «теоремы Шеннона» были доказаны безукоризненно и в достаточно общих предположениях. Наиболее законченный характер из работ этого направления имеет работа А. Я. Хинчина [6].

С значительно опережающим фундаментальные матемтические исследования развитием теории информации, осуществляемым исследователями более прикладного направления, советский читатель может с достаточной полнотой познакомиться по книгам [7], [8], [9] (в [7] включен перевод основной работы Шеннона [10], книга Гольдмана [9] вскоре должна выйти в русском переводе).

Существенные элементы теории информации для непрерывного случая, казалось бы более трудного, возникли до Шеннона. Логарифмическая мера количества информации, вполне аналогичная шенноновскому выражению $H(\xi)$, лежит в основе ассимптотических методов статистики, разработанных Фишером еще в 1921—1925 годах (см. [11], гл. 32—33*).

Еще ближе непосредственно к работам Шеннона стоят результаты В. А. Котельникова [12], полученные еще в 1933 году. Здесь была сформулирована фундаментальная идея спектральной теории передачи информации при помощи непрерывных сигналов, о которой я говорю подробнее в главе II этого доклада**.

ЛИТЕРАТУРА

1. *А. Н. Колмогоров.* Статистическая теория колебаний с непрерывным спектром. Материалы общего собрания АН СССР 1947 г.
2. *А. Н. Колмогоров.* Статистическая теория колебаний с непрерывным спектром. Юбилейный сборник АН СССР, т. I, 1947, 242—254.

---

* Общеизвестно и то обстоятельство, что выражение $H(\xi)$ формально тождественно с выражением энтропии в физике. Это совпадение я считаю вполне достаточным для оправдания наименования величины $H(\xi)$ и в теории информации «энтропией»: такие математические аналогии следует всегда подчеркивать, так как сосредоточение на них внимания содействует прогрессу науки. Но было бы преувеличением считать, что физические теории, связанные с понятием энтропии, уже содержат в себе в готовом виде элементы теории информации: первое использование выражений типа энтропии в качестве меры количества информации, которое мне известно, находится в только что упомянутых работах Фишера.

** Вторая глава настоящей брошюры содержит больше математических подробностей, чем устный доклад.

3. *А. Н. Колмогоров.* О некоторых асимптотических характеристиках вполне ограниченных метрических пространств. ДАН СССР, **108**, № 3, 1956, 385—388.
4. *А. Г. Витушкин.* К тринадцатой проблеме Гильберта. ДАН СССР, 95, 4, 1954, 701—704.
5. *А. Я. Хинчин.* Понятие энтропии в теории вероятностей. УМН, VIII, вып. 3. (55), 1953, 3—20.
6. *А. Я. Хинчин.* Об основных теоремах теории информации. УМН, XI, вып. 1 (67), 1956, 17—75.
7. «Теория передачи электрических сигналов при наличии помех». Сборник переводов. М., 1953.
8. *А. А. Харкевич.* Очерки общей теории связи. Гос. изд-во техн.-теор. лит-ры, М., 1955.
9. *Goldman Stanford.* Information theory. New York, 1953.
10. *C. E. Shannon a. W. Weaver.* The Mathematical Theory of Communications. Univ. of Illinois Press, 1949, 3—89 (есть русский перевод, см. [7]).
11. *Г. Крамер.* Математические методы статистики. М., 1948.
12. *В. А. Котельников.* Материалы к 1-му Всесоюзному съезду по вопросам реконструкции связи. 1933.

II. ПРИНЦИПЫ ПОСТРОЕНИЯ ТЕОРИИ В СЛУЧАЕ НЕПРЕРЫВНЫХ СООБЩЕНИЙ

Роль энтропии случайного объекта ξ, способного принимать значения $x_1, x_2, \ldots, x_n$ с вероятностями $p_1, p_2, \ldots, p_n$,

$$H(\xi) = -\sum_k p_k \log p_k$$

в теории информации и теории передачи сообщений при помощи дискретных сигналов можно считать достаточно выясненной. Далее я настаиваю на той идее, что основным понятием, допускающим обобщение на совершенно произвольные непрерывные сообщения и сигналы, является не непосредственно понятие энтропии, а понятие количества информации $I(\xi, \eta)$ в случайном объекте ξ относительно объекта η. В дискретном случае эта величина корректно вычисляется по известной формуле Шеннона *

$$I(\xi, \eta) = H(\eta) - MH(\eta \mid \xi).$$

Для конечномерных распределений, обладающих плотностью, величина $I(\xi, \eta)$, по Шеннону, определяется аналогичной формулой:

$$I(\xi, \eta) = h(\eta) - Mh(\eta \mid \xi),$$

где $h(\eta)$ есть «дифференциальная энтропия»

$$h(\eta) = -\int p_\eta(y) \log p_\eta(y)\, dy,$$

а $h(\eta \mid \xi)$ — аналогичным образом определяемая условная дифференциальная энтропия. Общеизвестно, что величина $h(\xi)$

---

* В кажущихся мне целесообразными обозначениях $H(\eta \mid x)$ есть условная энтропия η при $\xi = x$, а $MH(\eta \mid \xi)$ — математическое ожидание этой условной энтропии при переменном ξ.

не имеет непосредственного реального истолкования и даже не инвариантна по отношению к преобразованиям координат в пространстве x-ов. Для бесконечномерных распределений аналога выражения $h(\xi)$, вообще говоря, не существует.

В собственном смысле слова, энтропия объекта ξ с непрерывным распределением всегда бесконечна. Если непрерывные сигналы не могут тем не менее служить для передачи неограниченно большой информации, то только потому, что они всегда наблюдаются с ограниченной точностью. Естественно поэтому, задав точность наблюдения ε, определить соответствующую ей «ε — энтропию» $H_\varepsilon(\xi)$ объекта ξ. Это и было сделано Шенноном под названием «скорости создания сообщений». Хотя выбор для этой величины нового названия и не меняет существа дела, я решаюсь предложить такое переименование, подчеркивающее более широкий интерес понятия и его глубокую аналогию с обыкновенной точной энтропией. Укажу заранее на то, что, как указывается в § 3, для ε — энтропии сохраняется теорема об экстремальной роли нормального распределения (как в конечномерном, так и в бесконечномерном случае). Далее, в § 1—2, я даю, не претендуя на безусловную новизну, абстрактную формулировку определения и основных свойств величины $I(\xi, \eta)$ и обзор основной проблематики теории передачи сообщений Шеннона, В § 3—6 излагаются некоторые конкретные результаты, полученные советскими математиками в самое последнее время. Мне особенно хотелось бы подчеркнуть интерес исследований по асимптотическому поведению ε — энтропии при $\varepsilon \to 0$. Исследованные ранее случаи

$$H_\varepsilon(\xi) \backsim n \log \frac{1}{\varepsilon}, \quad \overline{H}_\varepsilon(\xi) \backsim 2W \log \frac{1}{\varepsilon},$$

где n — число измерений, а W — ширина полосы спектра, являются лишь очень частными случаями могущих здесь встретиться закономерностей. Для понимания открывающихся здесь перспектив может представлять интерес, изложенная в других терминах моя заметка [5].

§ 1. Количество информации в одном случайном объекте относительно другого

Пусть ξ и η являются случайными объектами с областями возможных значений X и Y,

$$P_\xi(A) = \mathrm{P}(\xi \in A), \quad P_\eta(B) = \mathrm{P}(\eta \in B),$$

соответствующие распределения вероятностей и

$$P_{\xi\eta}(C) = \mathrm{P}((\xi, \eta) \in C)$$

совместное распределение вероятностей объектов ξ и η. По определению количество информации в случайном объекте ξ относительно случайного объекта η дается формулой

$$I(\xi, \eta) = \int_X \int_Y P_{\xi\eta}(dx\,dy) \log \frac{P_{\xi\eta}(dx\,dy)}{P_\xi(dx)\,P_\eta(dy)}. \tag{1}$$

Точный смысл этой формулы требует некоторых пояснений, а сообщаемые далее общие свойства величины $I(\xi, \eta)$ справедливы лишь при некоторых теоретико-множественного характера ограничениях на распределения P_ξ, P_η и $P_{\xi\eta}$, но здесь я не буду на этом останавливаться. Во всяком случае, общая теория без больших затруднений может быть изложена таким образом, что она будет применима к случайным объектам ξ и η весьма общей природы (векторам, функциям, обобщенным функциям и т. п.). Определение (1) можно считать принадлежащим Шеннону, хотя он и ограничился случаем

$$P_\xi(A) = \int_A p_\xi(x)\,dx, \quad P_\eta(B) = \int_B p_\eta(y)\,dy,$$

$$P_{\xi\eta}(C) = \iint_C p_{\xi\eta}(x, y)\,dx\,dy,$$

когда (1) переходит в

$$I(\xi, \eta) = \int_X \int_Y p_{\xi\eta}(x, y) \log \frac{p_{\xi\eta}(x, y)}{p_\xi(x)\,p_\eta(y)}\,dx\,dy.$$

Иногда бывает полезно представить распределение $P_{\xi\eta}$ в виде

$$P_{\xi\eta}(C) = \iint_C a(x, y)\,P_\xi(dx)\,P_\eta(dy) + S(C), \tag{2}$$

где функция $S(C)$ сингулярна относительно произведения

$$P_\xi \times P_\eta.$$

Если сингулярная компонента S отсутствует, то формула

$$\alpha_{\xi, \eta} = a(\xi, \eta) \tag{3}$$

определяет однозначно с точностью до вероятности нуль случайную величину $\alpha_{\xi, \eta}$. Иногда бывает полезна сформулированная И. М. Гельфандом и А. М. Ягломом [1].

ТЕОРЕМА. Если $S(X \times Y) > 0$, то $I(\xi, \eta) = \infty$. Если $S(X \times Y) = 0$, то

$$I(\xi, \eta) = \int_X \int_Y a(x, y) \log a(x, y) P_\xi(dx) P_\eta(dy) =$$
$$= \int_X \int_Y \log a(x, y) P_{\xi\eta}(dx \cdot dy) = M \log \alpha_{\xi, \eta}. \quad (4)$$

Перечислим некоторые основные свойства величины $I(\xi, \eta)$.

I. $I(\xi, \eta) = I(\eta, \xi)$.

II. $I(\xi, \eta) \geqslant 0$; $I(\xi, \eta) = 0$ лишь в случае, если ξ и η независимы.

III. Если пары (ξ_1, η_1) и (ξ_2, η_2) независимы, то

$$I((\xi_1, \xi_2), (\eta_1, \eta_2)) = I(\xi_1, \eta_1) + I(\xi_2, \eta_2).$$

IV. $I((\xi, \eta), \zeta) \geqslant I(\eta, \zeta)$.

V. $I((\xi, \eta), \zeta) = I(\eta, \zeta)$
в том и только в том случае, если ξ, η, ζ есть марковская последовательность, т. е. если условное распределение ζ при фиксированных ξ и η зависит только от η.

По поводу свойства IV полезно заметить следующее.

В случае энтропии

$$H(\xi) = I(\xi, \xi)$$

кроме оценок энтропии пары (ξ, η) снизу

$$H(\xi, \eta) \geqslant H(\xi), \quad H(\xi, \eta) \geqslant H(\eta),$$

вытекающих из I и IV, имеется оценка сверху

$$H(\xi, \eta) \leqslant H(\xi) + H(\eta).$$

Для количества информации в ζ относительно пары (ξ, η) аналогичной оценки не существует: из

$$I(\xi, \zeta) = 0, \quad I(\eta, \zeta) = 0$$

еще не вытекает, как можно показать на элементарных примерах, равенство

$$I((\xi, \eta), \zeta) = 0.$$

Для дальнейшего отметим специально случай, когда ξ и η являются случайными векторами

$$\xi = (\xi_1, \ldots, \xi_m),$$
$$\eta = (\eta_1, \ldots, \eta_n) = (\xi_{m+1}, \ldots, \xi_{m+n})$$

и величины

$$\xi_1, \xi_2, \ldots, \xi_{m+n}$$

распределены нормально со вторыми центральными моментами

$$s_{ij} = M\,[(\xi_i - M\xi_i)(\xi_j - M\xi_j)].$$

Если детерминант

$$C = |\, s_{ij} \,|\; 1 \leqslant i,\, j \leqslant m + n$$

отличен от нуля, то как было подсчитано И. М. Гельфандом и А. М. Ягломом,

$$I(\xi, \eta) = \frac{1}{2} \log \frac{AB}{C}, \qquad (5)$$

где

$$A = |\, s_{ij} \,|_{1 \leqslant i,\, j \leqslant m}, \quad B = |\, s_{ij} \,|_{m < i,\, j \leqslant m+n}.$$

Часто целесообразнее, впрочем, другой подход к делу, применимый без ограничения $C > 0$. Как известно [2], после надлежащего линейного преобразования координат в пространствах X и Y все вторые моменты s_{ij}, кроме тех, для которых $i = j$, или $j = m + i$, обратятся в нуль. При таком выборе координат

$$I(\xi, \eta) = -\frac{1}{2} \log \sum_k [1 - r^2(\xi_k, \eta_k)], \qquad (6)$$

где суммирование берется по тем

$$k \leqslant \min(m, n),$$

для которых знаменатель в выражении коэффициента корреляции

$$r(\xi_k, \eta_k) = \frac{s_{k,\, m+k}}{\sqrt{s_{k,k} \cdot s_{m+k,\, m+k}}}$$

отличен от нуля.

§ 2. Абстрактное изложение основ теории Шеннона

Шеннон рассматривает передачу сообщений по схеме

$$\xi \to \eta \to \eta' \to \xi',$$

где «передающее устройство»

$$\eta \to \eta'$$

характеризуется условным распределением

$$P_{\eta'|\eta}(B' \mid y) = \mathrm{P}(\eta' \in B' \mid \eta = y)$$

«сигнала на выходе» η' при заданном «сигнале на входе» η и некоторыми ограничениями

$$P \in V.$$

распределения P_η входного сигнала. Операции «кодирования»

$$\xi \to \eta$$

и «декодирования»

$$\eta' \to \xi'$$

характеризуются условными распределениями

$$P_{\eta|\xi}(B/x) = \mathrm{P}(\eta \in B/\xi = x),$$

$$P_{\xi'|\eta'}(A'/y') = \mathrm{P}(\xi' \in A'/\eta' = y').$$

Основная задача Шеннона заключается в следующем. Заданы пространства X, X', Y, Y' возможных значений «сообщений на входе» ξ, «сообщений на выходе» ξ', сигналов на входе η и сигналов на выходе η', заданы характеристики передающего устройства, т. е. условные распределения $P_{\eta'|\eta}$, и класс V допустимых распределений P_η входного сигнала; заданы, наконец, распределение

$$P_\xi(A) = \mathrm{P}(\xi \in A)$$

сообщения на входе и «требования к точности передачи»

$$P_{\xi\xi'} \in W,$$

где W некоторый класс совместных распределений

$$P_{\xi\xi'}(C) = \mathrm{P}((\xi, \xi') \in C)$$

сообщения на входе и сообщения на выходе. Спрашивается: возможно ли, и в случае возможности — каким способом, задать правила кодирования и декодирования (т. е. условные распределения $P_{\eta|\xi}$ и $P_{\xi'|\eta'}$) таким образом, что, вычисляя распределение $P_{\xi\xi'}$, по распределениям P_ξ, $P_{\eta|\xi}$, $P_{\eta'|\eta}$, $P_{\xi'|\xi}$ в предположении, что последовательность

$$\xi, \eta, \eta', \xi'$$

марковская, получим

$$P_{\xi\xi'} \in W?$$

Определим вместе с Шенноном «пропускную способность» передающего устройства формулой

$$C = \sup_{P_\eta \in V} I(\eta, \eta')$$

и введем величину

$$H_W(\xi) = \inf_{P_{\xi\xi'} \in W} I(\xi, \xi'),$$

которую при расчете на единицу времени Шеннон называет «скоростью создания сообщений». Тогда, из свойства V § 1 сразу вытекает необходимое условие возможности передачи

$$H_W(\xi) \leqslant C. \tag{7}$$

Как уже отмечалось, несравненно глубже идея Шеннона, что в применении к длительной работе «каналов связи» условие (7) является в некотором смысле и при некоторых весьма широких условиях и «почти достаточным». С математической точки зрения здесь дело идет о доказательстве предельных теорем следующего типа. Предполагается, что пространства X, X', Y, Y', распределения P_ξ и $P_{\eta' \mid \eta}$, классы V и W, а следовательно, и величины C и $H_W(\xi)$ зависят от параметра T (играющего в применениях роль длительности работы передающего устройства). Требуется установить при некоторых весьма общего характера предположениях, что условие

$$\liminf_{T\to\infty} \frac{C^T}{H_W^T(\xi)} > 1 \tag{8}$$

достаточно для возможности передачи, удовлетворяющей поставленным выше условиям, при достаточно больших T. Естественно, что в такой постановке задача несколько расплывчата. Однако я умышленно избегнул здесь обращения к терминологии теории стационарных случайных процессов, так как можно получить в намеченном направлении не лишенные интереса результаты и без предположения стационарности [3].

Выводу предельных теорем указанного типа для дискретного случая посвящено много замечательных работ. Среди них особенно следует отметить уже называвшуюся работу А. Я. Хинчина [4].

Для общего непрерывного случая здесь, собственно, еще почти ничего не сделано. ~~Но для случая, когда ξ, η, η', ξ' являются случайными величинами (конечномерными или бесконечномерными, например, случайными функциями $\xi(t)$, $\eta(t)$, $\eta'(t)$, $\xi'(t)$ от времени t), подчиненными нормальному закону распределения, выполнение этой программы не должно встретить никаких существенных затруднений.~~

§ 3. Вычисление и оценка ε-энтропии в некоторых частных случаях

Если условие

$$P_{\xi, \xi'} \in W$$

выбрать в виде обязательности точного совпадения ξ и ξ'

$$P(\xi = \xi') = 1,$$

то

$$H_W(\xi) = H(\xi).$$

В соответствии с этим кажется естественным назвать в общем случае $H_W(\varepsilon)$ «энтропией случайного объекта ξ при точности воспроизведения W».

Предположим теперь, что пространство X' совпадает с X, т. е. исследуются способы приближенной передачи сообщения о положении точки $\xi \in X$ при помощи указания точки ξ' того же пространства X, и что в пространстве введено «расстояние» $\rho(x, x')$, удовлетворяющее обычным аксиомам «метрических пространств». Кажется естественным потребовать, чтобы

$$P\{\rho(\xi, \xi') \leqslant \varepsilon\} = 1, \qquad (W_\varepsilon^0)$$

или чтобы

$$M\rho^2(\xi, \xi') \leqslant \varepsilon^2. \qquad (W_\varepsilon)$$

Эти два вида «ε-энтропии» распределения P_ξ будем обозначать

$$H_{W_\varepsilon^0}(\xi) = H_\varepsilon^0(\xi),$$

$$H_{W_\varepsilon}(\xi) = H_\varepsilon(\xi).$$

Что касается ε-энтропии $H_\varepsilon^0(\xi)$, то я хочу здесь лишь отметить некоторые оценки для

$$H_\varepsilon^0(X) = \sup_{P_\xi} H_\varepsilon^0(\xi),$$

где верхняя грань берется по всем распределениям вероятностей P_ξ на пространстве X. При $\varepsilon = 0$, как известно,

$$H_0^0(X) = \sup_{P_\xi} H(\xi) = \log N_X,$$

где N_X есть число элементов множества X. При $\varepsilon > 0$

$$\log N_X^c(2\varepsilon) \leqslant H_\varepsilon^0(X) \leqslant \log N_X^a(\varepsilon),$$

где $N_X^a(\varepsilon)$ и N_X^c являются характеристиками пространства X, которые введены в моей заметке [5]. Ассимптотические свойства при $\varepsilon \to 0$ функций $N_X(\varepsilon)$, изученные для ряда конкретных пространств X в [5], являются интересными аналогами излагаемых далее свойств ассимптотического поведения функции $H_\varepsilon(\xi)$.

Обратимся теперь к ε-энтропии $H_\varepsilon(\xi)$. Если X есть n-мерное эвклидово пространство и

$$P_\xi(A) = \int_A p_\xi(x)\, dx_1\, dx_2 \ldots dx_n,$$

то, по крайней мере в случае достаточно гладкой функции $p_\xi(x)$, имеет место хорошо известная формула

$$H_\varepsilon(\xi) = n \log \frac{1}{\varepsilon} + [h(\xi) - n \log \sqrt{2\pi e}] + o(1), \qquad (9)$$

где

$$h(\xi) = -\int_X p_\xi(x) \log p_\xi(x)\, dx_1\, dx_2 \ldots dx_n$$

«дифференциальная энтропия», введенная уже в первых работах Шеннона. Таким образом, ассимптотическое поведение $H_\varepsilon(\xi)$ в случае достаточно гладких непрерывных распределений в n-мерном пространстве определяется **в первую очередь** размерностью пространства и лишь в виде второго члена в выражение $H_\varepsilon(\xi)$ входит дифференциальная энтропия $h(\xi)$.

Естественно ожидать, что для типичных распределений в бесконечномерных пространствах рост $H_\varepsilon(\xi)$ при $\varepsilon \to 0$ будет существенно более быстрым. В виде простейшего примера рассмотрим случайную функцию Винера $\xi(t)$, определенную для $0 \leqslant t \leqslant 1$, с нормально распределенными независимыми приращениями

$$\Delta\xi = \xi(t + \Delta t) - \xi(t),$$

для которой

$$\xi(0) = 0, \quad M\Delta\xi = 0, \quad M(\Delta\xi)^2 = \Delta t.$$

А. М. Яглом нашел, что в этом случае в метрике L^2

$$H_\varepsilon(\xi) = \frac{4}{\pi} \cdot \frac{1}{\varepsilon^2} + o\left(\frac{1}{\varepsilon^2}\right). \qquad (10)$$

Более общим образом для процесса Маркова диффузного типа на отрезке времени $t_0 \leqslant t \leqslant t_1$ с

$$M\Delta\xi = A(t, \xi(t))\, \Delta t + o(\Delta t); \qquad M(\Delta\xi)^2 = B(t, \xi(t))\, \Delta t + o(\Delta t)$$

можно при некоторых естественных допущениях получить формулу

$$H_\varepsilon(\xi) = \frac{4}{\pi} \chi \cdot \frac{1}{\varepsilon^2} + o\left(\frac{1}{\varepsilon^2}\right), \qquad (11)$$

где

$$\chi(\xi) = \int_{t_0}^{t_1} MB(t, \xi(t))\, dt.$$

Для случая нормального распределения в n-мерном эвклидовом, или в гильбертовом пространстве ε-энтропия H_ε может быть вычислена точно; n-мерный вектор ξ после надлежащего ортогонального преобразования координат приобретает вид

$$\xi = (\xi_1, \xi_2, \ldots, \xi_n),$$

где координаты ξ_k взаимно независимы и распределены нормально. При заданном ε параметр θ определяется из уравнения

$$\varepsilon^2 = \sum \min(\theta^2, D\xi_k)$$

и, в случае нормально распределенного ξ,

$$H_\varepsilon(\xi) = \frac{1}{2} \sum_{D\xi_k > \theta^2} \log \frac{D\xi_k}{\theta^2}. \tag{12}$$

Аппроксимирующий вектор

$$\xi' = (\xi'_1, \xi'_2, \ldots, \xi'_n)$$

следует выбирать так, что при $D\xi_k \leqslant \theta^2$

$$\xi'_k = 0,$$

а при $D\xi_k > \theta$

$$\xi_k = \xi'_k + \Delta_k, \quad D\Delta_k = \theta^2, \quad D\xi'_k = D\xi_k - \theta^2$$

и векторы ξ_k и Δ_k взаимно независимы. Бесконечномерный случай ничем не отличается от конечномерного.

Наконец, весьма существенно, что **максимальное значение $H_\varepsilon(\xi)$ для вектора ξ** (n-мерного, или бесконечномерного) **при заданных вторых центральных моментах достигается в случае нормального распределения**. Этот результат может быть получен непосредственно или из следующего предложения М. С. Пинскера (ср. [6]):

ТЕОРЕМА. Пусть заданы положительно определенная симметрическая матрица величин s_{ij}, $0 \leqslant i, j \leqslant m+n$, и распределение P_ξ вектора

$$\xi = (\xi_1, \xi_2, \ldots, \xi_m),$$

центральные вторые моменты которого равны s_{ij} (при $0 \leqslant i, j \leqslant m$). Пусть условие W на совместное распределение $P_{\xi, \xi'}$ вектора ξ и вектора

$$\xi' = (\xi_{m+1}, \xi_{m+2}, \ldots, \xi_{m+n})$$

заключается в том, что центральные вторые моменты величин

$$\xi_1, \xi_2, \ldots, \xi_{m+n}$$

равны s_{ij} (при $0 \leqslant i, j \leqslant m+n$). Тогда

$$H_W(\xi) \leqslant \frac{1}{2} \log \frac{AB}{C}. \qquad (13)$$

Обозначения в формуле (13) соответствуют изложению § 1. Из сопоставления со сказанным в § 1 видно, что неравенство (13) превращается в равенство в случае нормального распределения P_ξ.

Принципы решения вариационных задач, возникающих при вычислении «скорости создания сообщений», были указаны довольно давно Шенноном. В [7] Шеннон и Уивер пишут: «К сожалению, эти формальные решения в частных случаях трудно численно оценить, и поэтому ценность их представляется небольшой» *. По существу, однако, многие задачи такого рода, как видно из сказанного выше, достаточно просты. Возможно, что медленное развитие исследований в этом направлении связано с недостаточным пониманием того обстоятельства, что в типичных случаях решения вариационных задач оказываются очень часто вырожденными; например, в разобранной выше задаче вычисления $H_\varepsilon(\xi)$ для нормально распределенного вектора ξ в n-мерном случае вектор ξ' часто оказывается не n-мерным, а лишь k-мерным с $k < n$, в бесконечномерном же случае вектор ξ' оказывается всегда конечномерным.

§ 4. Количество информации и скорость создания сообщений в случае стационарных процессов

Рассмотрим два стационарных и стационарно связанных процесса

$$\xi(t), \eta(t), \qquad -\infty < t < \infty.$$

Будем обозначать через ξ_T и η_T отрезки процессов ξ и η за время $0 < t \leqslant T$ и через ξ_- и η_- течение процессов ξ и η на отрицательной полуоси $-\infty < t \leqslant 0$. Задать пару (ξ, η) стационарно связанных процессов ξ и η это значит задать инвариантное по отношению к сдвигам вдоль оси t распределение вероятностей $P_{\xi\eta}$ в пространстве пар функций $\{x(t), y(t)\}$. Если зафиксировать ξ_-, то из распределения $P_{\xi\eta}$ возникает услов-

* § 28, стр. 79 русского перевода.

ное распределение

$$P_{\xi_T, \eta|\xi_-}(C\,|x_-) = P\{(\xi_T, \eta)\in C\,|\,\xi_- = x_-\}.$$

При помощи этого распределения в соответствии с § 1 вычисляется условное количество информации

$$I(\xi_T, \eta\,|\,x).$$

Если математическое ожидание

$$MI(\xi_T, \eta\,|\,\xi_-)$$

конечно при каком либо $T > 0$, то оно конечно и при всех других $T > 0$ и

$$MI(\xi_T, \eta\,|\,\xi_-) = T\vec{I}(\xi, t).$$

Величину $\vec{I}(\xi, \eta)$ естественно назвать «скоростью создания информации о процессе η при наблюдении процесса ξ». Если процесс ξ с полной точностью экстраполируется из прошлого в будущее, то

$$\vec{I}(\xi, \eta) = 0.$$

Так будет, в частности, если процесс ξ имеет ограниченный спектр. Вообще говоря, не обязательно имеет место равенство

$$\vec{I}(\xi, \eta) = \vec{I}(\eta, \xi). \tag{14}$$

Однако при довольно широких условиях «регулярности» процесса * ξ имеет место равенство

$$\vec{I}(\xi, \eta) = \bar{I}(\xi, \eta),$$

где

$$\bar{I}(\xi, \eta) = \lim_{T\to\infty} \frac{1}{T} I(\xi_T, \eta_T).$$

Так как $I(\xi_T, \eta_T) = I(\eta_T, \xi_T)$, то всегда

$$\bar{I}(\xi, \eta) = \bar{I}(\eta, \xi)$$

и, следовательно, в случае справедливости обоих равенств $\vec{I}(\xi, \eta) = \bar{I}(\xi, \eta)$ и $\vec{I}(\eta, \xi) = \bar{I}(\eta, \xi)$, имеет место равенство (14). Пусть теперь W есть некоторый класс совместных распределений $P_{\xi\xi'}$ двух стационарных и стационарно связанных процессов ξ и ξ'. Естественно назвать величину

$$\vec{H}_W(\xi)\ \inf_{P_{\xi\xi'}\in W} \vec{I}(\xi', \xi)$$

* Примечание на стр. 28.

«скоростью создания сообщений в процессе ξ при точности воспроизведения W». При некоторых предположениях регулярности процесса ξ и для некоторых естественных типов условий W можно доказать, что

$$\vec{H}_W(\xi) = \overline{H}_W(\xi),$$

где

$$\overline{H}_W = \inf_{P_{\xi\xi'} \in W} \overline{I}(\xi, \xi').$$

§ 5. Вычисление и оценка количества информации и скорости создания сообщений по спектру

Для случая, когда распределение $P_{\xi\eta}$ нормально и хотя бы один из процессов ξ или η регулярен, имеет место предложенная М. С. Пинскером [9] формула

$$\overline{I}(\xi, \eta) = -\frac{1}{4\pi}\int_{-\infty}^{+\infty} \log[1 - r^2(\lambda)]\, d\lambda,$$

где (15)

$$r^2(\lambda) = \frac{|f_{\xi\eta}(\lambda)|^2}{f_{\xi\xi}(\lambda) f_{\eta\eta}(\lambda)},$$

а $f_{\xi\xi}$, $f_{\xi\eta}$, $f_{\eta\eta}$ — спектральные плотности. В случае процессов с дискретным временем t для дифференциальной энтропии нормального процесса на единицу времени

$$\overline{h}(\xi) = \lim_{T \to \infty} \frac{1}{T} h(\xi_1, \xi_2, \ldots, \xi_T)$$

известно выражение

$$\overline{h}(\xi) = \log(2\pi\sqrt{e}) + \frac{1}{2\pi}\int_{-\pi}^{+\pi} \log f_{\xi\xi}(\lambda)\, d\lambda. \qquad (16)$$

Однако, в случае непрерывного времени и неограниченного спектра никакого аналога выражения $\overline{h}(\xi)$ не существует и формула Пинскера требует самостоятельного вывода.

Естественно характеризовать точность воспроизведения стационарного процесса ξ при помощи стационарного и стационар-

* Регулярность процесса здесь и далее обозначает, грубо говоря, что отрезки процесса, соответствующие двум достаточно удаленным друг от друга отрезкам оси t, почти независимы. В случае гауссовских процессов здесь применимо хорошо известное определение регулярности, введенное в моей работе [8].

но связанного с ξ процесса ξ' величиной

$$\sigma^2 = M[\xi(t) - \xi'(t)]^2$$

и в случае условия W вида

$$\sigma^2 \leqslant \varepsilon^2$$

называть величину

$$\overline{H}_\varepsilon(\xi) = \overline{H}_W(\xi)$$

ε — энтропией на единицу времени процесса ξ, а в предположении, что

$$\overrightarrow{H}_W(\xi) = \overline{H}_W(\xi)$$

скоростью создания сообщений в процессе ξ при средней точности передачи ε. Из соответствующего предложения для конечномерных распределений (см. § 3) можно вывести, что при заданной спектральной плотности $f_{\xi,\xi}(\lambda)$ величина $\overline{H}_\varepsilon(\xi)$ достигает максимума в случае нормального процесса ξ. В нормальном случае величина $\overline{H}_\varepsilon(\xi)$ может быть легко вычислена по спектральной плотности $f_{\xi\xi}(\lambda)$ вполне аналогично тому, как это было объяснено в § 3 в применении к величине $H_\varepsilon(\xi)$ для n-мерных распределений. Параметр θ определяется из уравнения

$$\varepsilon^2 = \int_{-\infty}^{+\infty} \min(\theta^2, f_{\xi\xi}(\lambda))\, d\lambda. \tag{17}$$

При помощи этого параметра величина $\overline{H}_\varepsilon(\xi)$ находится по формуле

$$\overline{H}(\xi) = \frac{1}{2} \int_{f_{\xi\xi}(\lambda) > \theta^2} \log \frac{f_{\xi\xi}(\lambda)}{\theta^2}\, d\lambda. \tag{18}$$

Практический интерес представляют спектральные плотности вида (рис. 2), которые хорошо аппроксимируются функцией

$$\varphi(\lambda) = \begin{cases} a^2 & \text{при } A \leqslant |\lambda| \leqslant A + W, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Легко подсчитать, что в этом случае п р и н е с л и ш к о м м а л ы х ε приближенно для нормального процесса

$$\begin{aligned} \theta^2 &\sim \frac{\varepsilon^2}{2W} \\ \overline{H}_\varepsilon(\xi) &\sim W \log \frac{2Wa^2}{\varepsilon^2}. \end{aligned} \tag{19}$$

Формула (19), конечно, есть не что иное, как хорошо известная формула Шеннона

$$R = W \log \frac{Q}{N}. \qquad (20)$$

Принципиальная новизна заключается здесь, однако, в том, что теперь мы видим почему и в каких пределах (при не

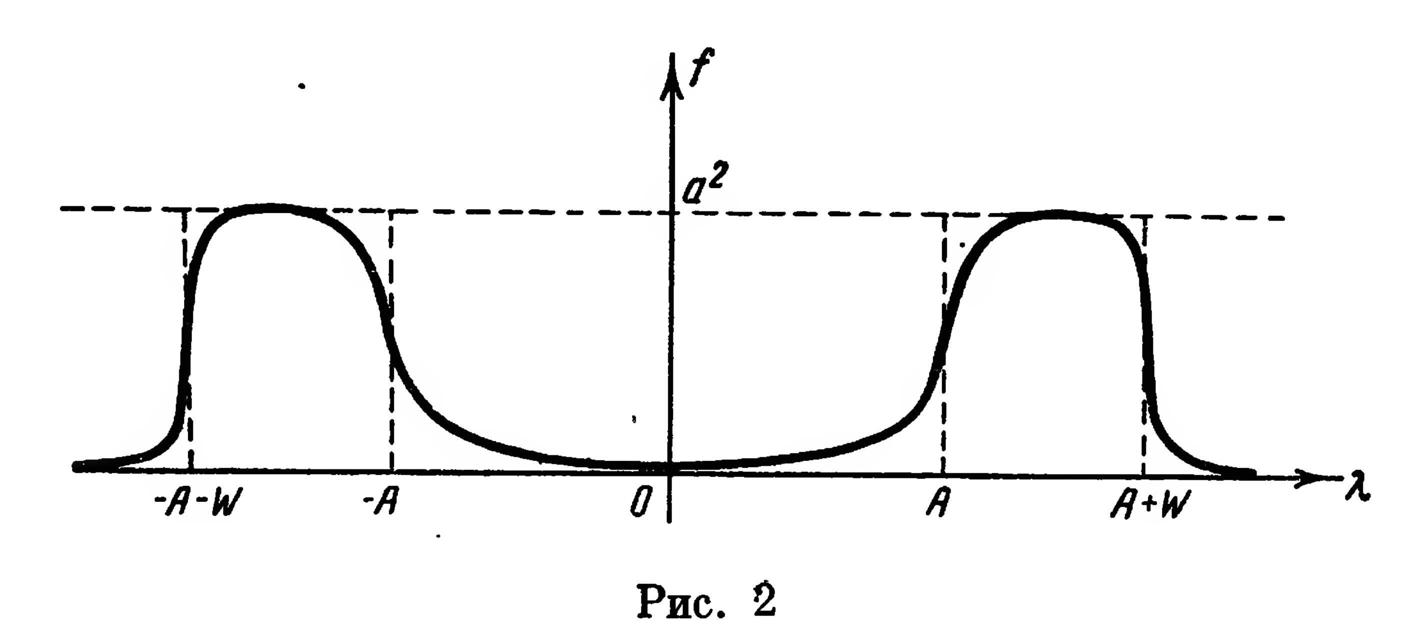


Рис. 2

слишком малых ε) эта формула может быть применима к процессам с неограниченным спектром, а таковы все реально интересующие нас в теории передачи сообщений процессы. Записывая (19) в виде

$$\overline{H}_{\varepsilon}(\xi) \approx 2W \log (a \sqrt{W}) \log \frac{1}{\varepsilon} \qquad (21)$$

и сравнивая с (9), мы видим, что двойная ширина $2W$ используемой полосы частот играет роль числа измерений. Эта идея эквивалентности двойной ширины полосы частот числу измерений, приходящихся, в некотором смысле слова, на единицу времени, была, повидимому, впервые высказана В. А. Котельниковым (см. [10]). В обоснование этой идеи Котельников указывал на то обстоятельство, что функция, спектр которой помещается в полосе ширины $2W$, однозначно определяется значениями функции в точках

$$\ldots, -\frac{2}{2W}, -\frac{1}{2W}, 0, \frac{1}{2W}, \frac{2}{2W}, \ldots, \frac{k}{2W}, \ldots$$

Эта же аргументация сохранена и у Шеннона, использующего полученные таким образом представления и для вывода формулы (20). Так как функция с ограниченным спектром всегда сингулярна в смысле [8] и наблюдение такой функции вообще не связано со стационарным притоком новой информации, то смысл такого рода аргументации оставался не вполне ясным, так что приведенный здесь новый вывод **приближенной** формулы (21) представляется мне не лишенным интереса.

При малых ε для любой нормально распределенной регулярной случайной функции рост $\overline{H}_\varepsilon(\xi)$ с убыванием ε происходит существенно быстрее, чем это получалось бы по формуле (21). В частности, если $f_{\xi\xi}(\lambda)$ при $\lambda \to \infty$ имеет порядок $\lambda^{-\beta}$, то $\overline{H}_\varepsilon(\xi)$ имеет порядок $\varepsilon^{-2/(\beta-1)}$.

§ 6. Вычисление и оценка пропускной способности в некоторых частных случаях

Сначала рассмотрим случай, когда входной сигнал является m-мерным вектором

$$\eta = (\eta_1, \ldots, \eta_m),$$

а выходной сигнал — n-мерным вектором

$$\eta' = (\eta'_1, \ldots, \eta'_n).$$

Как уже говорилось, передающее устройство $\eta \to \eta'$ характеризуется условным распределением вероятностей $P_{\eta'|\eta}$ и некоторыми ограничениями на входной сигнал η. Предположим, что зависимость η' от η имеет характер линейной нормальной корреляции, т. е., что

$$\eta' = A\eta + \zeta, \tag{22}$$

где оператор A линеен, а вектор ζ независим от η и подчинен n-мерному гауссовскому распределению. Что касается условий на входной сигнал, то допустим, что они имеют вид

$$MQ(\eta) \leqslant \varepsilon^2, \tag{23}$$

где Q есть некоторая положительно определенная квадратичная форма от координат $\eta_1, \ldots, \eta_m$. При помощи надлежащего выбора линейных преобразований координат в пространствах Y и Y' общий случай сводится к случаю, когда

$$Q(\eta) = \Sigma\eta_i^2, \tag{24}$$

а зависимость (22) записывается в виде

$$\left.\begin{aligned} \eta'_i &= a_i\eta_i + \zeta_i, \quad a_i \neq 0, \quad \text{при } 1 \leqslant i \leqslant k, \\ \eta'_i &= \zeta_i, \qquad\qquad\qquad\quad \text{при } i > k, \end{aligned}\right\} \tag{25}$$

где $k = \min(m, n)$.

В сделанных предположениях пропускная способность, т. е. верхняя грань C количества информации $I(\eta, \eta')$, совместимая с наложенными условиями, легко находится. Она достигается,

если η_i — взаимно независимые случайные гауссовские величины

$$\left.\begin{aligned} D\eta_i &= \theta^2 - \frac{D\zeta_i}{a_i^2}, \text{ если } i \leqslant k \text{ и } a_i^2\theta^2 > D\zeta_i, \\ D\eta_i &= 0 \text{ при остальных } i, \end{aligned}\right\} \tag{26}$$

где константа θ^2, как легко видеть, однозначно определяется требованиями (26) и

$$\sum_i D\eta_i \leqslant \varepsilon^2. \tag{27}$$

Соответствующее значение $C = I(\eta, \eta')$ есть (см. § 3)

$$C = \frac{1}{2} \sum_{a_i^2\theta^2 > D\zeta_i} \log \frac{a_i^2\theta^2}{D\zeta_i}. \tag{28}$$

Аналогично дело обстоит в случае стационарных линейных каналов связи с «гауссовскими шумами», т. е. в случае

$$\eta' = A\eta(t) + \zeta(t), \tag{29}$$

где A линейный оператор, а $\zeta(t)$ независимая от $\eta(t)$ гауссовская стационарная случайная функция. Как известно, соответствующие спектральные плотности связаны соотношением

$$f_{\eta'\eta'}(\lambda) = a^2(\lambda) f_{\eta\eta}(\lambda) + f_{\zeta\zeta}(\lambda). \tag{30}$$

Условие, ограничивающее мощность входного канала, зададим в форме

$$\int_{-\infty}^{+\infty} f_{\eta\eta}(\lambda)\, d\lambda \leqslant \varepsilon^2. \tag{31}$$

Формальное решение задачи вполне аналогично решению рассмотренной выше конечномерной задачи. Наибольшая пропускная способность канала $\bar{C}$ достигается, если

$$\left.\begin{aligned} f_{\eta\eta}(\lambda) &= \theta^2 - a^2(\lambda)/f_{\zeta\zeta}(\lambda) \text{ при } a^2(\lambda)/f_{\zeta\zeta}(\lambda) > \theta^2, \\ f_{\eta\eta}(\lambda) &= 0 \text{ при } a^2(\lambda)/f_{\zeta\zeta}(\lambda) \leqslant \theta^2, \end{aligned}\right\} \tag{32}$$

Соответствующее значение $\bar{I}(\eta, \eta')$ по формуле (15) равно

$$\bar{C} = \frac{1}{4\pi} \int_{a^2(\lambda)\theta^2 > f_{\zeta\zeta}(\lambda)} \log \frac{a^2(\lambda)\,\theta^2}{f_{\zeta\zeta}(\lambda)}\, d\lambda. \tag{33}$$

$\frac{f_{\zeta\zeta}(\lambda)}{a^2(\lambda)}$ при $a^2(\lambda)\theta^2 > f_{\zeta\zeta}(\lambda)$

$a^2(\lambda)\theta^2 \leq f_{\zeta\zeta}(\lambda)$

Следует тут же заметить, что в практически интересных случаях вычисленная по формулам (32) спектральная плотность $f_{\eta\eta}(\lambda)$ обращается в нуль за пределами ограниченного участка оси λ и поэтому является спектральной плотностью сингулярного процесса. Процесс $\eta'(t)$ будет при этом (в естественном предположении регулярности процесса $\zeta(t)$) смешанным — т. е. состоять из суммы сингулярной компоненты $A\eta$ и регулярной компоненты ζ. Скорость $\vec{I}(\eta', \eta)$ создания информации о процессе η при наблюдении процесса η' в таком случае вовсе не выражается формулой $\vec{I}(\eta', \eta) = \overline{C}$, а равна нулю (см. § 5).

При помощи некоторых дополнительных рассмотрений можно, однако, показать, что, вводя в виде входного сигнала $\eta(t)$ регулярные процессы, которые в принципе практически реализуемы и действительно несут возникающую во времени информацию, можно получить скорость создания информации $\vec{I}(\eta', \eta) = \vec{I}(\eta, \eta')$, сколь угодно близкую к $\overline{C}$. Это окончательно оправдывает формулу (33).

ЛИТЕРАТУРА

1. *А. Н. Колмогоров, А. М. Яглом, И. М. Гельфанд.* Количество информации и энтропия для непрерывных распределений. Доклад на Третьем всесоюзном математическом съезде. 1956.
2. *А. М. Обухов.* Нормальная корреляция векторов. Ученые записки МГУ, вып. 45, 1940, 73—92.
3. *Р. М. Розенблат.* Труды Третьего всесоюзного математического съезда, т. II, 1956, 132—133.
4. *А. Я. Хинчин.* «Успехи математических наук», 11, вып. 1, 1956, 17—75.
5. *А. Н. Колмогоров.* ДАН СССР, 108, 1956, 385—388.
6. *М. С. Пинскер.* Труды Третьего всесоюзного математического съезда, т. I, 125.
7. *C. E. Shannon a. W. Weaver.* The Mathematical Theory of Communications. Univ. of Illinois Press, 1949, 3—89 (есть русский перевод в сборнике переводов «Теория передачи электрических сигналов при наличии помех». М., 1953).
8. *А. Н. Колмогоров.* Бюллетень МГУ, т. 2, вып. 6, 1941.
9. *М. С. Пинскер.* ДАН СССР, 98, 1954, 213—216.
10. *В. А. Котельников.* Материалы к 1-му Всесоюзному съезду по вопросам реконструкции связи. 1933.

Сдано в набор 10/X 1956 г. Подп. к печ. 13/X 1956 г.
Формат бум. 60×92 1/16. Печ. л. 2,25. Уч.-издат. л. 2
Тип. заказ № 919. Т-09350. Тираж 3300

Бесплатно

Издательство АН СССР
Москва Б-64, Подсосенский пер., д. 21

2-я типография Издательства АН СССР
Москва Г-99, Шубинский пер. д. 10

Бесплатно